

ALGORITHM FOR FAST DISK BASED TEXT MINING

ABSTRACT OF THE DISCLOSURE

Methods and apparatus, including computer systems and program products, for executing a query, for example, a query for a document similar to another document. In one general aspect, the techniques feature a method of executing a query for at least one document similar to a specified document. That method includes receiving the query; forming a reduced query document based on ranks of terms in the specified document; generating a modified query based on the query and the reduced query document; executing the modified query on a data repository to generate a set of results; and, providing a result to a user interface.